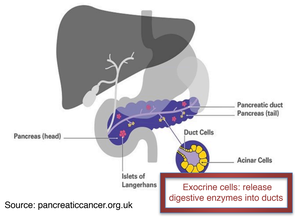
Intron Retention is a Robust Marker of Intertumoral Heterogeneity in Pancreatic Ductal Adenocarcinoma

Daniel J. Tan[1,4], **Mithun Mitra**[1,2*], **Alec M. Chiu**[3] and **Hilary A. Coller**[1,2,3*]

[1]Department of Molecular, Cell and Developmental Biology, University of California, Los Angeles; [2]Department of Biological Chemistry, David Geffen School of Medicine, University of California, Los Angeles; [3]Bioinformatics Interdepartmental Program, University of California, Los Angeles; [4]Present address: Department of Biomedical Informatics, Harvard Medical School; *These authors jointly supervised this work

1 What is Pancreatic Ductal Adenocarcinoma? (PDAC)

Anatomy of the Pancreas



Source: pancreaticcancer.org.uk

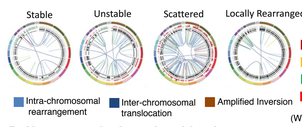
- PDAC is a cancer of the exocrine pancreas cells
- PDAC constitutes about >90% of pancreatic cancers.
- PDAC is the 4th leading cause of cancer-related deaths in the US. By 2030, PDAC is anticipated to become the 2nd leading cause of cancer-related deaths after lung cancer.
- Overall 5-year survival rate is 8% for all stages combined.
- Majority of patients are diagnosed at a late stage and for this group the 5-year survival is 3%.
- PDAC tumors are quite chemoresistant due to broad heterogeneity of genetic mutations and dense stroma.

2 Goals of This Work

1. Better define the molecular nature of inter-tumoral heterogeneity in PDAC patients in terms of variation in alternative splicing.
2. Identify molecular subtypes of PDAC through alternative splicing data and correlating them with clinical outcomes.
3. Identify novel splicing targets specific to new molecular subtypes.
4. Identify differentially spliced events and investigate their effects.
5. Identify RNA-binding proteins that could be responsible for differential splicing.

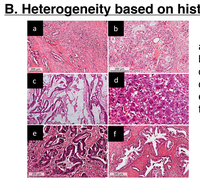
3 Previous studies in PDAC heterogeneity

A. Heterogeneity based on frequency and distribution of structural rearrangements



(Waddell, et al., Nature Medicine, 2016)

B. Heterogeneity based on histology



a) Classical
b) Adenosquamous
c) Colloid
d) Undifferentiated
e) Intestinal Variant
f) Micropapillary

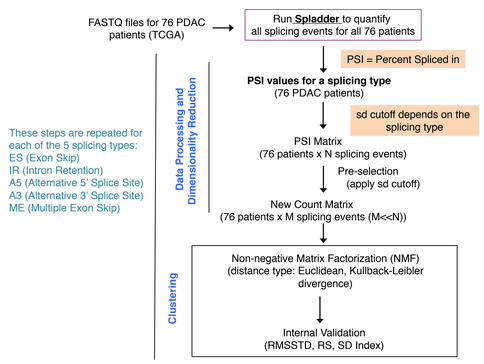
(Cross, et al., Pathobiology, 2018)

C. Heterogeneity based on gene expression

	Collisson et al. (Nature Medicine, 2011)	Moffitt et al. (Nature Genetics, 2015)	Bailey et al. (Nature, 2016)
Sample	microdissected PDA (n=27) and GSE15471	Primary PDAC tumors (n=145) and other samples	96 tumors (>40% epithelial content)
Dataset type	Microarray	Microarray (RNA-seq on 15 primary tumors)	RNA-seq
Subtypes	Exocrine like, Classical, Quasi-Mesenchymal	basal, classical (did consider tumor stroma)	squamous, ADEX, pancreatic progenitor, immunogenic
Clustering method	Non negative matrix factorization (NMF)	Non negative matrix factorization (NMF)	Non negative matrix factorization (NMF)
Gene signature	62 genes	25 genes each for basal and classical	10 gene expression programs
Survival comparison	Classical better than QM (p=0.038)	Classical better than basal (p=0.007)	Worst survival for squamous (p=0.03)

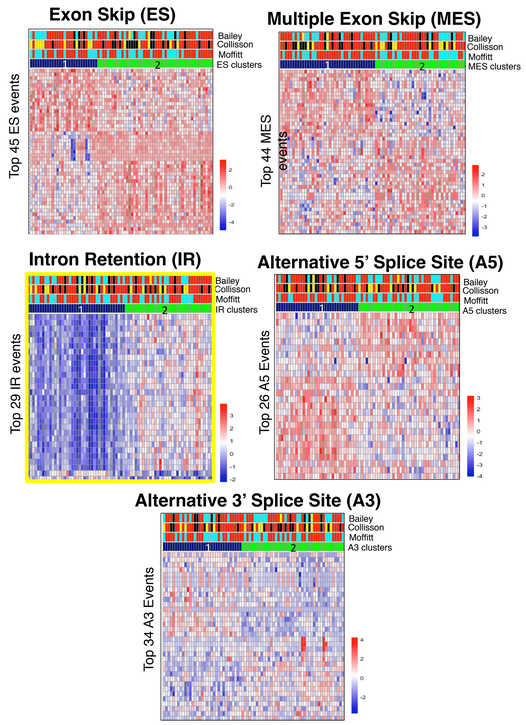
(Moffitt et al., Nature Genetics, 2015)
(Bailey et al., Nature, 2011)
(Collisson et al., Nature Medicine, 2011)

4 Pipeline for NMF Clustering based on Alternative Splicing

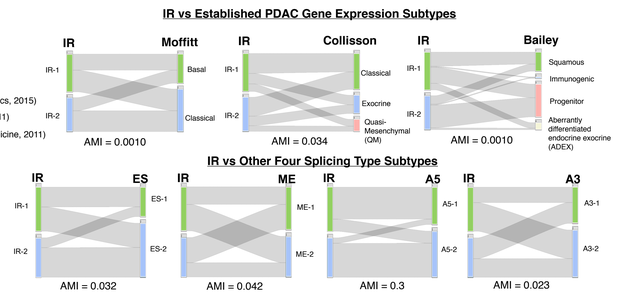


These steps are repeated for each of the 5 splicing types:
ES (Exon Skip)
IR (Intron Retention)
A5 (Alternative 5' Splice Site)
A3 (Alternative 3' Splice Site)
ME (Multiple Exon Skip)

5 IR shows optimal clustering out of the five splicing types, based on measures of cluster compactness and separation

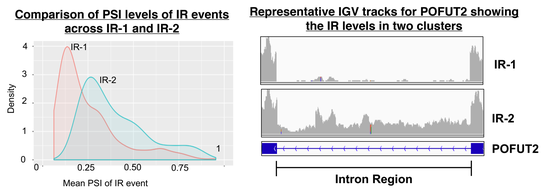


6 IR clusters show little to no concordance with the established PDAC gene expression subtypes or with the stratifications of the four other splicing types

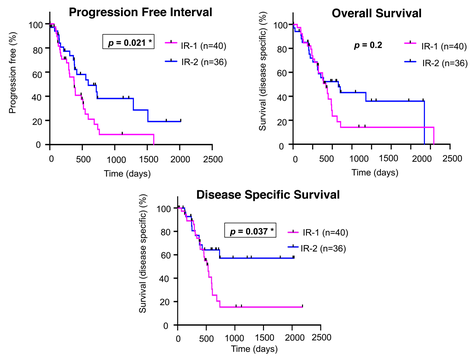


*AMI (Adjusted Mutual Information) is a measure of similarity between two assignments (0 to 1), with 1 being the highest score.

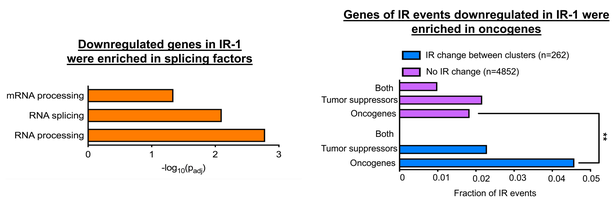
7 Higher levels of IR are asymmetrically and globally skewed towards IR-2



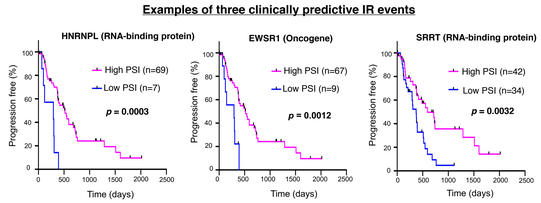
8 IR-1 has significantly poorer clinical outcomes compared to IR-2



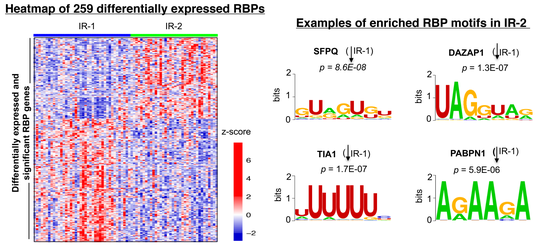
9 GO Analysis of genes undergoing differential intron retention



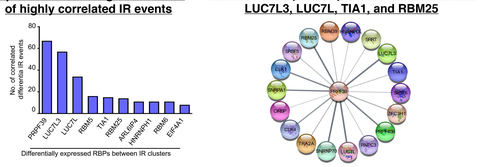
10 20 IR events were independently predictive of clinical outcome



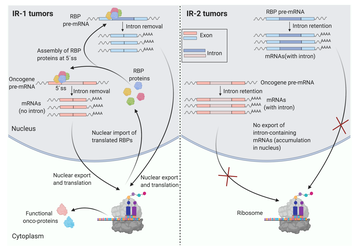
11 258 RNA-binding proteins (RBPs) were differentially expressed, and motifs for 35 RBPs were enriched



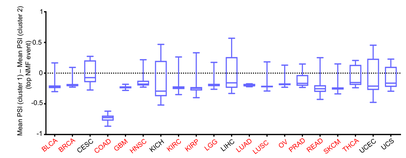
12 Expression of 25 RBPs correlated highly with PSI levels of 139 / 262 IR events; many of these RBPs interact with each other at the protein level



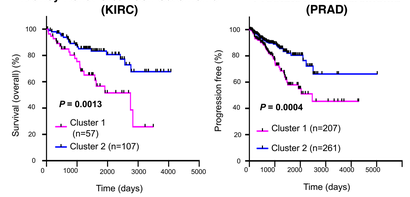
13 IR may be a mechanism for regulating expression of oncogenes via nuclear detention



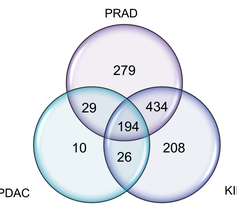
14 IR is similarly asymmetrically distributed between tumor clusters in 15/20 other cancers



15 IR was found to be predictive of clinical outcome in two other cancers: KIRC and PRAD



16 194 common RBPs were differentially expressed between PDAC, PRAD, and KIRC



- The 194 shared RBPs include 23 / 25 of the highly correlated RBPs found in PDAC
- This suggests similar features in the RBP networks among these cancer types
- IR, and specific RBP networks could be important contributors to tumor-tumor variation in various cancer types

CONCLUSIONS

1. Intron Retention produces the most statistically robust clustering out of five different splicing and displays differences in clinical outcomes between clusters
2. Intron Retention explains an independent dimension of heterogeneity compared to gene expression and is the only asymmetrically distributed AS event; it occurs at lower levels in the low clinical outcome PDAC patient population.
3. Genes undergoing differential IR were significantly enriched for splicing factors and oncogenes
4. Splicing differences between the two IR clusters can potentially be explained by differentially expressed splicing factors and RNA-binding proteins.
5. IR may be leading to differential nuclear detention of oncogenic transcripts between IR-1 and IR-2.
6. IR was also found to be predictive of clinical outcome in KIRC and PRAD.

Acknowledgements

Members of Coller Lab

We thank Dr. Eleazar Eskin for use of his high-performance computing nodes.

We thank the Hoffman2 Shared Cluster support staff for their help in resolving technical issues.

Questions or Comments?

Contact:
dtan@hms.harvard.edu

or

Join Slack Channel:
https://bit.ly/2zLfweK